Appendix B. NASS/CDS Sample Design

The crashes investigated in NASS/CDS are a probability sample of all police-reported crashes in the United States. Each such crash that occurs within a CDS team's area has a chance of being included in the sample. This design makes it possible to compute not only national estimates but also probable errors associated with those estimates. Many other features of the design have a significant impact on CDS data analysis, the most important of which are highlighted in this appendix.

The selection of sample crashes for CDS is accomplished in stages. The first stage is the selection of geographic areas called primary sample units (PSUs). Each PSU is composed of a large city, a county, or a group of contiguous counties. The United States was divided into 1,195 PSUs. The PSUs were then grouped into 12 categories described by geographic region and degree of urbanization. Two PSUs were selected from each category with probability proportional to its 1983 population. These 24 PSUs are the first stage in the selection of CDS sample crashes.

If every crash in each of the 24 PSUs were investigated, a national estimate could be obtained by weighting each crash in the PSU by the inverse of the probability of selection of the PSU. For example, if a sample PSU had 1 chance in 40 of being selected, then each crash from the PSU would be weighted by a factor of 40. This is called the first-stage expansion factor.

It is not practical to investigate every crash in each sample PSU, so additional stages of sampling are performed. The police agencies in a PSU are categorized by the number and type of police crash reports they process. Sample police agencies are then selected randomly from each category. The fraction of the agencies selected increases as the number and severity of crashes reported by the agency increases. This is called the second-stage expansion factor.

The final stage of sampling is the selection of crashes from all crashes reported in the sample police agencies. A simple random selection of all reported towaway crashes would result in a large percentage of sample crashes with property damage and few injuries, since these constitute such a large fraction of all crashes. This type of sample would not be effective in providing the detailed and accurate information needed for the mitigation of crash consequences. Rather, a substantial sample of serious injury crashes is needed for NASS/CDS.

The procedure used to obtain the desired sample by type and severity of crashes is an unequal probability selection. This required listing police crash reports in categories defined by most severe police-reported injury to an occupant of a towed CDS applicable motor vehicle, disposition of the injured, and model year of the towed CDS applicable motor vehicle. A weighting factor was assigned to crashes in each category to increase or decrease the probability of selection. A random selection was made from the total crashes listed in all categories. In addition to the probabilities of selection varying by type of crash, other factors affected the selection probabilities at this stage, such as the number of crashes listed, the date and time of the crash, and the police agencies from which the

crash was listed. The result was that each sampled crash from a PSU has a unique selection probability.

The inverse of this probability is called the third-stage expansion factor. If each sample crash in a PSU is multiplied by its second- and third-stage expansion factors, an unbiased estimate of the total number of crashes in the PSU is obtained. To produce the national estimates, the PSU level estimates are inflated by the first-stage expansion factor. Thus, the national expansion factor is the product of the first-, second-, and third-stage expansion factors.

The national estimates equal the inverse of the product of the probability of the PSUs being selected, the probability of the police agencies being selected, and the probability of the crash being selected for that day. Since the number of crashes in the sample is predetermined, the national estimate for each crash within a stratum is different. To account for this bias, a ratio weight was developed. The ratio weight is the national estimate multiplied by a ratio factor. For each stratum, this ratio factor is equal to the total number of crashes listed in all of the police jurisdictions (sampled and non-sampled) divided by the number of crashes selected. There are instances where very few or no crashes are listed. To account for this, the similar PSUs were grouped together, based on the stratum from which they were originally selected.